

Towards jointly learned control policies and image recovery for distributed aperture telescopes

Justin Fletcher^a and Peter Sadowski^a

^aUniversity of Hawai'i at Manoa

ABSTRACT

Extended object imaging of resident space objects is foundational to space domain awareness. Large aperture optical systems can capture spatially resolved imagery of objects in low Earth orbit but are infeasible for objects at higher altitudes. In this work we explore the application of distributed aperture optical systems to extended object imaging at distances beyond low Earth orbit using fully differentiable physical models. Distributed aperture systems are a cost-effective design for large aperture telescopes, but entail a sequential control problem to correct for aberrations induced by phase differences between the spatially distal subapertures. We provide a fully differentiable, joint formulation of this control problem and the associated image recovery task, and train a model to maximize reconstruction quality from an ensemble of focal plane images. We measure the quality of the recovered images, and position these results relative to the recovery quality achieved by monolithic telescopes.

Keywords: differentiable physics, neural image recovery, interferometry, space domain awareness

1. INTRODUCTION

Timely, accurate, and detailed assessments of satellite location and state underpin space traffic management (STM), and enable safe access to space for commercial and defense applications. While satellites may be detected and tracked using affordable small-diameter telescopes,^{1,2} ground-based extended imaging of satellites is only feasible using large telescopes; further, optical diffraction complicates extended image formation for satellites beyond low Earth orbit (LEO), even with correction of wavefront errors due to atmospheric turbulence. To image more distant objects, more resolving power is required, which in turn requires larger primary apertures under traditional telescope design approaches. Unfortunately, telescope cost increases rapidly with aperture diameter,³ which limits the scalability of traditional designs. This cost dynamic and the proliferation of resident space objects (RSOs) conspire to create a challenging resource allocation problem for commercial or national entities seeking to enhance space domain awareness using extended object imaging.

A new large telescope design concept, the ExoLife Finder (ELF), could enable more cost-effective extended satellite imaging via distributed aperture synthesis.⁴ However, the increased complexity of this design introduces technical challenges related to subaperture phase control and image recovery. In this work we model the image formation, phase control, and image recovery problems *jointly*, proposing an end-to-end, differentiable model of the entire distributed aperture optical system (Fig. 1). This approach parallels similar approaches being explored in the nascent field of metasurface optics.

The principle contributions of this work are: (1) a novel, differentiable formulation of distributed aperture image formation enabling task-gradient driven selection of secondary mirror actuation; (2) a neural network image recovery model; and (3) experimental results demonstrating better image reconstruction with learnable articulations, including analysis of the impact of both the number of images formed and the size of the optical baseline of the system on aggregate recovery quality.

J.F.: jfletch@hawaii.edu

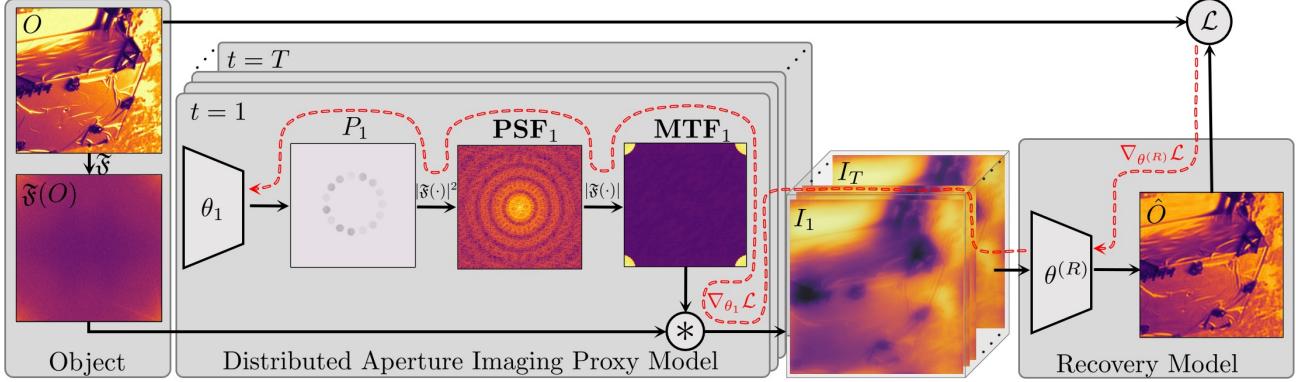


Figure 1: We model a distributed aperture optical system as a differentiable, parameterized function approximator, and jointly optimize a control policy and image recovery algorithm.

2. RELATED WORKS

2.1 Large Telescopes and Space Domain Awareness

Space domain awareness (SDA) is the discipline concerned with building and maintaining awareness of RSOs, including their location, operational status, normal behaviour, and the intent of their operators. Recent work has demonstrated image processing methods to extract information relevant to SDA from extended images of satellites.^{5–9} The information provided by these methods is valuable to the SDA community, but is available only from comparatively few large-aperture telescopes.



Figure 2: A render of the ExoLife Finder telescope design. The primary aperture is structured as an annulus of circular subapertures, which reduces moving mass and cost.

Distributed aperture telescopes have the potential to be much cheaper at larger scales, but must overcome optical aberrations due to lower structural rigidity.³ The ELF design, shown in Fig. 3, includes articulating secondary apertures, which allow for active compensation through tip, tilt, and piston articulation control. Modern articulation controllers operate at kilohertz update rates when run open-loop. Throughout this work, when we discuss subaperture articulation it is this selection of tip, tilt, and piston commands for the secondary apertures of each subaperture to which we refer. Because we select articulation commands during training, the

resulting control policy is open-loop. The encouraging results provided here provide motivation and a baseline for comparison of closed-loop control policies, which we will explore in future work.

2.2 Differentiable Physics

Our architecture and approach are analogous to recent work in metasurface thin lens imaging.¹⁰ That work entails the optimization of a differentiable proxy model for a nano-optic lens design – also expressed as a polynomial, as in Sec. 3 – and subsequent non-blind, learned deconvolution for image recovery. Both that work and ours fit within the emerging field of differentiable physics;¹¹ Tseng et al. apply the techniques of that field to the problem of metasurface nano-optic design and production, while we apply them to the problem of extremely large telescope articulation design. In both cases, the physics of propagating light is modeled using a differentiable proxy that, once realized (i.e., as a lens, or articulations of a distributed aperture telescope), shapes interference to yield features that are informative to a trainable image recovery model. We adopt a similar image recovery model architecture, employing a convolution neural network comprising a multi-scale feature extractor¹² with skip connections¹³ followed by a small-scale U-net.¹⁴

3. APPROACH

Our objective is to construct an optical system that enables performant image recovery, so we begin with image formation. Adopting the Fraunhofer diffraction model for far-field objects,¹⁵ and neglecting both focal plane noise and atmospheric effects*, we may model the image plane, I , as

$$\mathfrak{F}(I) = \mathbf{MTF} \circ \mathfrak{F}(O) \quad (1)$$

where \mathbf{MTF} is the modulation transfer function (MTF) of the optical system and O is the object plane (i.e., perfect) image.

This formulation elucidates the motivating intuition for this work: insofar as we can manipulate the MTF of the optical system, we can select spatial frequencies to emphasize in the resultant image. Given some imaging domain (e.g., one represented by a dataset), we might then choose to emphasize features that are, in expectation, relevant to solving a task (recognition, superresolution, etc.) in the context of that domain. Thus, by leveraging domain-specific and task-specific information – as represented by a dataset and task loss – we may select articulations to produce spatial frequency features that enhance task performance in that domain. Further, consider solving this problem sequentially through time: one may design an ensemble of sequential articulations (and therefore features) to maximize that ensemble's collective diversity of features, conditioned upon some recovery algorithm that maps the ensemble of images to a single reconstruction. This raises a challenging, but intriguing, question: how might we *jointly* design an articulation policy and a recovery algorithm, conditioned upon the image domain, to maximize the quality of the recovered image?

3.1 A Differentiable Model of Optical Systems

We confront this challenge by modeling both the recovery algorithm *and the optical system* as differentiable, parameterized estimators which we optimize jointly to minimize reconstruction end-to-end error. This approach requires that the optical system be differentiable with respect to the articulations of the tip, tilt, and piston actuation of each subaperture; as such, we build our model beginning at the MTF and work backward to a set of parameters corresponding to these articulation commands. The MTF of an optical system is the real component of the optical transfer function (OTF), which is the Fourier transform of the point spread function (PSF). The PSF, in turn, is given by the squared absolute value of the Fourier transform of the pupil function.¹⁶ Summarizing,

$$\mathbf{MTF} = |\mathbf{OTF}| = |\mathfrak{F}(\mathbf{PSF})| = \left| \mathfrak{F} \left(|\mathfrak{F}(P)|^2 \right) \right| \quad (2)$$

*We also assume a static target, though this is likely a reasonable assumption for the articulation frequencies and integration periods with which the work is concerned.

where **OTF**, **PSF**, and P are the OTF, PSF and pupil function, respectively. Both the absolute value function and the Fourier transform are supported by several leading autodifferentiation frameworks. This leaves only the pupil function, which we must now construct in terms of the tip, tilt, and piston articulations of each of the subapertures.

To build a differentiable pupil function corresponding to the ELF design, we must construct an annulus of circular subapertures. For this purpose, we model each subaperture, P_n as a radially positioned generalized Gaussian distribution (GGD),

$$P_n(u, v \mid \mu_u, \mu_v, \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-[(u-\mu_u)^2 + (v-\mu_v)^2/\alpha]^\beta}, \quad (3)$$

where μ_u and μ_v are the spatial coordinates of the subaperture centroid in the pupil plane. Given a sufficiently large β constant, P_n forms a differentiable approximation to a circular mask with the radius of the mask controlled by the GGD scale parameter, α . Because α is a dimensionless scale constant, we must also convert it to the units of the pupil plane (i.e., meters) to correctly scale the subapertures. To map from α to the subaperture radius, we observe that the edge of the simulated aperture corresponds to the point at which the density of the probability density function (PDF) modeled by the GGD takes on an arbitrarily small value, which we denote ϵ . Analyzing the one-dimensional, positive semi-definite case at the origin, we consider the point at which the PDF assumes this arbitrarily small value,

$$\epsilon = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{[x/\alpha]^\beta} \iff \epsilon = e^{[x/\alpha]^\beta} \iff \log_\beta(-\ln \epsilon) = \frac{x}{\alpha} \iff x = \log_\beta(-\ln \epsilon) \alpha. \quad (4)$$

We note that Eq. 4 is insensitive to the particular choice of ϵ , so long as $\epsilon \ll \beta$ and $\epsilon \ll 1$, both of which are necessary preconditions for this analysis. Thus, we may now express the extent of an aperture in the pupil function, modeled by a GGD, in the units over which that pupil function is defined.

We now have the tools to model a single subaperture without articulations. Modeling articulation of a given subaperture is comparatively straightforward, because we note that the GGD subaperture model has a density of approximately zero for all $x > \log_\beta(-\ln \epsilon) \alpha$. As such, we may model the desired articulations as a two dimensional articulation plane, A , given by

$$A_n(u, v \mid \mu_u, \mu_v, \theta_n^{(\text{tip})}, \theta_n^{(\text{tilt})}, \theta_n^{(\text{piston})}) = (u - \mu_u)\theta_n^{(\text{tip})} + (v - \mu_v)\theta_n^{(\text{tilt})} + \theta_n^{(\text{piston})}, \quad (5)$$

where $\theta_n^{(\text{tip})}$, $\theta_n^{(\text{tilt})}$, and $\theta_n^{(\text{piston})}$, are parameters specifying the tip, tilt, and piston of subaperture n , respectively. We represent an articulated aperture as the product of Eq. 3 and Eq. 5.

For single-aperture optical systems, the prior model is sufficient. For distributed aperture systems, we must also arrange the articulated apertures in the pupil plane to construct the composite pupil function. Computing the displacement of each aperture radially around the annulus is straightforward. However, we must also compute the subaperture radius, which is less so. Fixing the minimum displacement between subaperture edges, the outer radius of the annulus, and the number of subapertures constrains the subaperture radius. We follow this convention, which implies that variations in total optical baseline (e.g., Sec. 4.3) cause variations in the subaperture radius as well. We determine that the subaperture radius, r_s , is given by

$$r_s = \frac{r_a \sin\left(\frac{\pi}{N}\right) - \frac{r_d}{2}}{1 + \sin\left(\frac{\pi}{N}\right)}, \quad (6)$$

where r_d is the minimum displacement between subapertures, r_a is the outer radius of the annulus, and N is the number of subapertures.

Combining Eqs. 1-6 we may express the entire pupil function as a single parameterized differentiable function. The resulting expression is cumbersome and opaque; it is rendered more clearly as pseudocode, which is given

in Algorithm 1. Neglecting immutable parameterizations, which are made explicit in Algorithm 1, we can write the composite pupil function as

$$P(u, v \mid \theta) = \sum_{n=1}^N P_n(u, v) A(u, v \mid \theta_n) \quad (7)$$

where θ_n represents the articulation parameters of subaperture n , and $\theta = \{\theta_n \mid n = 1, 2, \dots, N\}$.

Algorithm 1 Simulate the image formed by an monochromatic, articulated distributed aperture optical system.

```

Require:  $N \in \mathbb{Z}$                                      ▷ The number of subapertures to simulate.
Require:  $q \in \mathbb{Z}$                                      ▷ The spatial quantization constant.
Require:  $\beta \in \mathbb{R}$                                      ▷ The GGD shape parameter.
Require:  $r_a \in \mathbb{R}$                                      ▷ The outer radius of the annulus of subapertures, in meters.
Require:  $r_d \in \mathbb{R}$                                      ▷ The minimum separation between subaperture edges, in meters.
Require:  $\lambda \in \mathbb{R}$                                      ▷ Monochromatic light wavelength, in microns.
Require:  $\text{FOV} \in \mathbb{R}$                                      ▷ Aperture field of view, in arcseconds.
Ensure:  $\theta^{(\text{tip})} \in \mathbb{R}^N$                                ▷ A set of  $N$  tip parameters, in microns per meter.
Ensure:  $\theta^{(\text{tilt})} \in \mathbb{R}^N$                                ▷ A set of  $N$  tilt parameters, in microns per meter.
Ensure:  $\theta^{(\text{piston})} \in \mathbb{R}^N$                              ▷ A set of  $N$  piston parameters, in microns.
Ensure:  $O \in \mathbb{R}^{q \times q}$                                      ▷ A  $q \times q$  object plane image.

1:  $r_s \leftarrow (r_a \sin(\frac{\pi}{N}) - \frac{r_d}{2}) / (1 + \sin(\frac{\pi}{N}))$            ▷ Compute the subaperture radius using Eq. 6.
2:  $\alpha \leftarrow r_s / \log_\beta(-\ln e^{-16})$                                          ▷ Map the  $\alpha$  to the GGD domain using Eq. 4.
3:  $E = \lambda q / (4.848 * \text{FOV})$                                          ▷ Compute the pupil extent
4: for  $n \in \{1, 2, \dots, N\}$  do                                              ▷ Iterate over each of  $N$  subapertures.
5:    $\mu_u \leftarrow (r_a - r_s) \cos(\frac{2\pi n}{N})$                                 ▷ Map radial to Cartesian coordinates.
6:    $\mu_v \leftarrow (r_a - r_s) \sin(\frac{2\pi n}{N})$                                 ▷ Map radial to Cartesian coordinates.
7:   for  $(u, v) \in \{\frac{-E}{2}, \dots, \frac{E}{2}\} \times \{\frac{-E}{2}, \dots, \frac{E}{2}\}$  do    ▷ Iterate over the quantized the pupil plane.
8:      $P_n(u, v) \leftarrow \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-[(u-\mu_u)^2 + (v-\mu_v)^2]^\beta/\alpha}$  ▷ Compute the GGD density using Eq. 3.
9:      $A_n(u, v) \leftarrow (u - \mu_u)\theta_n^{(\text{tip})} + (v - \mu_v)\theta_n^{(\text{tilt})} + \theta_n^{(\text{piston})}$  ▷ Apply  $\theta$  using Eq. 5.
10:    end for                                                               ▷ Add  $P_n A_n$  to  $P$ .
11:     $P \leftarrow P + P_n A_n$ 
12:  end for                                                               ▷ Use Eq. 1 and Eq. 2 to form an image of the object.
13:   $I \leftarrow \mathfrak{F} \left( \left| \mathfrak{F} \left( |P|^2 \right) \right| \circ \mathfrak{F}(O) \right)$ 
```

So far, we have only modeled the formation of a single image. To construct a sequence of images from the same object, we apply Eq. 7 to a sequence of parameterizations, $\theta = \{\theta_1, \theta_2, \dots, \theta_T\}$. This approach neglects image plane variations that would exist in practice; we are adopting it to establish an upper bound on performance against which future work may compare. In effect, the sequence of articulations represented by θ constitutes a *static control policy* that also corresponds to a trainable feature extractor. This feature extractor may be realized *physically* as directed optical interference in an optical system.

Thus, we arrive at a fully differentiable model of the optical system, including its articulation control policy. This model maps an object plane image to an ensemble of image plane images, modeling image formation. Now, given these images, we must combine them to recover an estimate of the original object plane image. How can we ensure that the recovery algorithm is able to use the features produced by the control policy?

3.2 Image Recovery

We approach image recovery as an autoencoding problem. Classical blind and non-blind image recovery techniques abound; we adopt a differentiable approach because back-propagating gradients through the recovery algorithm is critical to the broader task of joint control and restoration. It is the gradients generated at the input to the recovery algorithm that will serve as the task gradients for the differentiable optical system defined

in the preceding section. By propagating these gradients back into the optical system, we allow credit for failure and success in the recovered image to flow back to the articulation choices from which those outcomes are derived. Further, while many domain-specific image recovery techniques enjoy widespread use, we wish to design an approach in which any domain-specific (i.e., target-specific) information can be learned end-to-end.

For the specific choice of learned recovery approach, we take inspiration from recent work in metasurface optics.¹⁰ To minimize impediments to comparisons between the techniques, we adopt a close variant of the reconstruction approach used in that work. We hope to remove incidental details that might otherwise obscure the analogy between metasurface design for near-field sensing and our task-informed apodization approach to far-field sensing. The model used in this work is a down-sampling, multi-scale feature extractor followed by an up-sampling feature fusion network, with skip connection throughout. We omit the PSF propagation layer used in Tseng et al. and parameterize the number of filters used throughout the network, but otherwise implement the same architecture. The input to our network is the stack of images, one for each simulated articulation, output by our optical model; the output is a single reconstructed image, which we denote \hat{O} to emphasize that it is an estimate of the object image.

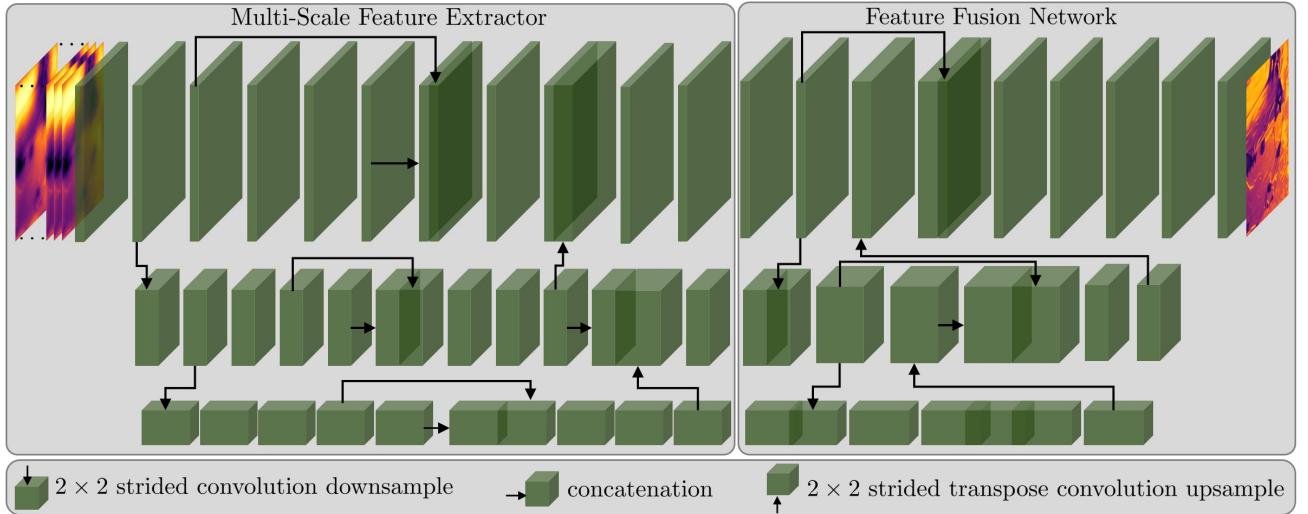


Figure 3: Our image recovery algorithm is a convolutional neural network comprising a multi-scale feature extractor, followed by feature fusion. In our implementation, filter depth is set throughout the network to be a multiple of a provided filter scale, with different depths at different levels of down-sampling.

We minimize the log of the pixel-wise mean-squared error (MSE) between a batch of reconstructed images and a batch of object plane images,

$$\mathcal{L} = \log \frac{1}{B} \sum_{b=1}^B \frac{1}{q^2} \sum_{p=1}^{q^2} \left(O_p^{(b)} - \hat{O}_p^{(b)} \right)^2, \quad (8)$$

where B is the batch size, q is the spatial quantization of the scene, and $O_p^{(b)}$ is the p -th pixel of b -th image in the batch. We minimize \mathcal{L} using adaptive momentum¹⁷ stochastic gradient descent. Image recovery performance is quantified using the ratio of the MSE of an image formed by a monolithic telescope to the MSE of a reconstructed image collected using a distribute aperture telescope with the same optical baseline, $\text{MSE}_{\text{m/d}}$. This choice of metric ensures that measured performance is not sensitive to the varying content of randomly-chosen crops from the dataset, and also allows for meaningful comparisons between datasets of dissimilar image complexity.

3.3 Datasets

The proposed end-to-end image recovery and articulation control solution is necessarily trained on finite data, and will exploit statistical properties of the training dataset. In this work, we use the SPEED+ lightbox dataset.¹⁸

This dataset comprises real images taken of a model satellite in many poses under a variety of lighting conditions. Similar variations of pose and lighting are present in ground-based extended imagery of satellites. SPEED+ is representative of a space-based satellite imaging scenario, and as such includes some features (e.g., Earth) that would not be present in a ground-based imaging scenario. These features only complicate the task, and do not meaningfully change the conclusions drawn in this work. The dataset includes only one satellite model. In Several examples from the dataset are shown in Fig. 4.



Figure 4: Four randomly chosen examples from the SPEED+ lightbox dataset, which is used in this work.

The portion of the SPEED+ dataset used in this work comprises 6066 images, each 1920×1200 . We randomly divide the dataset into training and validation sets of 5,392 and 674 images, respectively. During training and evaluation, images are randomly cropped to 512×512 . All results reported in Sec. 4 are computed using the held-out validation set.

3.4 Implementation

All code for this project is implemented in TensorFlow 2.2 using graph mode.¹⁹ Code for this project can be found at https://github.com/justinfletcher/differentiable_apodization. All experiments were conducted on Nvidia A100s. Each model was trained for 3,000 epochs.

4. RESULTS

In this section, we present the results of three experiments, each designed to probe features of the approach described in Sec. 3. We first assess the importance of end-to-end differentiability and static policy learning, respectively; then evaluate recovery performance across a variety of optical baselines (i.e., aperture outer diameters). Hyperparameter configurations are described in Table 1. Experimental results are reported in terms of validation set recovery performance as measured by the ratio of the simulated monolithic aperture MSE to the simulated DA MSE, $\text{MSE}_{\text{m/d}}$. A metric value of 1.0 indicates that the modeled DA system achieved a final image quality equivalent to that of a filled aperture of the same outer diameter, while metric values greater than 1.0 correspond to superior reconstruction. Because filled aperture telescopes cost considerably more due to their larger moving mass, equivalent or greater imaging performance by a distributed aperture system may motivate the adoption of that design.

4.1 Image Recovery and the Impact of Differentiability

The principal hypothesis of this work is that an end-to-end differentiable model of an optical and image recovery system conveys advantages in image recovery performance. We evaluate this hypothesis directly by fixing the parameters of the optical model. Each articulation parameter is initialized with a small displacement, in microns for piston and microns/meter for tip and tilt, which are drawn from a Gaussian with 0 mean and 0.001 standard deviation; these initial articulations are much smaller than would exist in practice. This leaves only the image recovery model parameters trainable. By stopping the gradients after the recovery model we allow the system to learn recovery, but prevent credit assignment back to the articulations which remain constant during training. This disentangles the effects of the recovery algorithm, which otherwise may be providing all relevant information about the object, from the joint effects of both the learned articulations and the recovery algorithm. The results of this experiment are shown in Fig. 6 and are tabulated in Table 2.

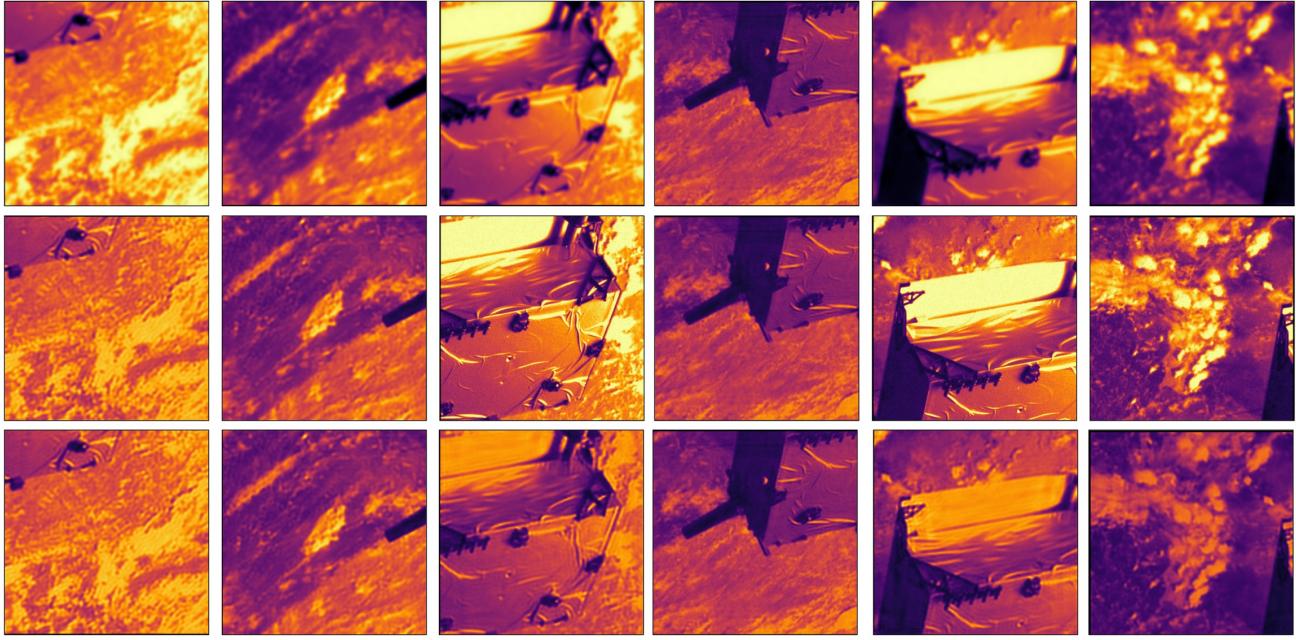


Figure 5: We report qualitatively successful image recovery, represented here by six random crops of a validation set image. To assist in recovery quality assessment, we position each the object plane image (**middle**) adjacent to the image formed by a monolithic aperture telescope (**top**) and a recovered image (**bottom**).

Parameter	Values	Experiment Section
Mutable Articulation	[True, False]	4.1
Ensemble Size	[1, 2, 4, 8]	4.2
Outer Diameter	[1.5, 2.5 , 3.5] meters	4.3
Monochromatic Wavelength	500 microns	-
Spatial Quantization & Crop Size	512	-
Filter Scale	4	-
Batch Size	32	-
Learning Rate	0.0001	-
Number of Subapertures	15	-
β	16	-

Table 1: The experiments conducted in this work vary several hyperparameters, while leaving others fixed. This identifies the hyperparameters used, and identifies the experiments in which they are varied. The default value of all hyperparameters are noted in bold.

Examining Fig. 6, we find support for the hypothesis that our differentiable optical model improves image recovery. Mutable articulations, which correspond to the presence of a differentiable optical model, improve end-to-end recovery beyond one standard deviation over five repetitions of the experiment. As shown in Table 2, we also observe that the maximum recovery performance achieved in any single trial is considerably higher when the control policy can be learned jointly with the recovery algorithm. Crucially, we report that learned end-to-end control and recovery enables distributed aperture imaging that matches or exceeds monolithic aperture imaging quality, as measured by MSE, while learned recovery alone does not.

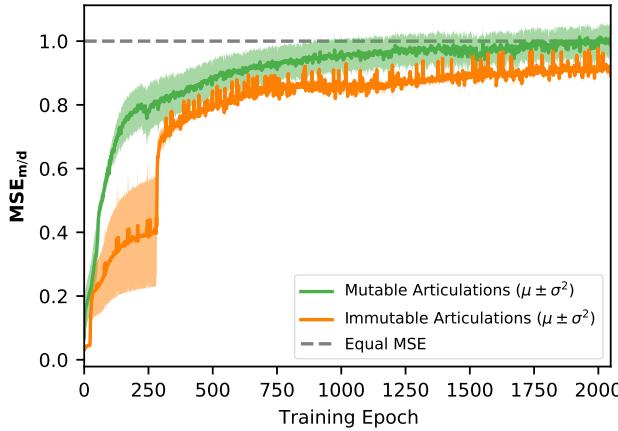


Figure 6: Differentiable, learned static control polices improve end-to-end validation set image recovery performance when used in conjunction with a learned image recovery model, relative to the same model with immutable static control. Bold lines indicate the 16-epoch moving average of the mean $MSE_{m/d}$ over four experiment repetitions, while the shaded regions denote one standard deviation of $MSE_{m/d}$ over the same repetitions.

Mutable Articulations	$MSE_{m/d}(\mu \pm \sigma^2)$	$MSE_{m/d}(\max)$
True	1.0132 ± 0.095	1.1624
False	0.9782 ± 0.0740	0.9280

Table 2: Quantitative measurements of validation set image recovery performance support the hypothesis that jointly learned control and recovery enhances performance. We report the mean and standard deviation $MSE_{m/d}$ achieved at the highest-performance epoch, as measured on the validation set. We also include the maximum $MSE_{m/d}$ achieved by each treatment, computed over all repetitions and all training epochs.

4.2 Ensemble Size

Next, we test the hypothesis that jointly training an ensemble of articulations (i.e., a control policy) enables the development of diverse, task-specific MTFs which aid the recovery model by providing richer input representation of the object. This is non-obvious. For instance, it could be the case that performance is always maximized by minimizing the phase errors induced by the articulations, which would imply that future work should seek only to phase subapertures, rather than to attempt controlled apodization. We investigate this intuition by comparing models trained with different ensemble sizes: 1, 2, 4, and 8. By including an experimental configuration of an ensemble size of 1, we ablate the effect of sequential exposures entirely. The results of this experiment are shown in Fig. 7 and Table 3.

We report that recovery performance improves with increasing ensemble size, and observe a significant degradation of performance for single-image ensembles. Again, we find that removing an essential component of the proposed approach – in this case optimization of a ensemble of articulation commands – prevents the modeled distributed aperture system from achieving image quality comparable to a monolithic aperture system of the same size.

4.3 Optical Baseline

Finally, we present preliminary experiments exploring the impact that the size of the simulated telescope systems has on image recovery quality. In this experiment we vary the outer diameter (i.e., optical baseline) of the simulated optical systems during training. We emphasize here that both the distributed aperture and monolithic systems are varied, which means that $MSE_{m/d} > 1.0$ implies that the distributed aperture image reconstruction achieved a lower MSE than a monolithic image formed by an aperture of the same diameter.

Relative MSE performance separates neatly according to optical baseline, as shown in Fig. 8, but not linearly: $MSE_{m/d}$ is lower at 3.5 m than 2.5 m. This result may arise the fact that, as the outer diameter is increased, the area of a monolithic aperture grows more quickly than the total area of a distributed aperture system. Alternatively, it may be related to the precision with which the distributed aperture MFT can be shaped as the

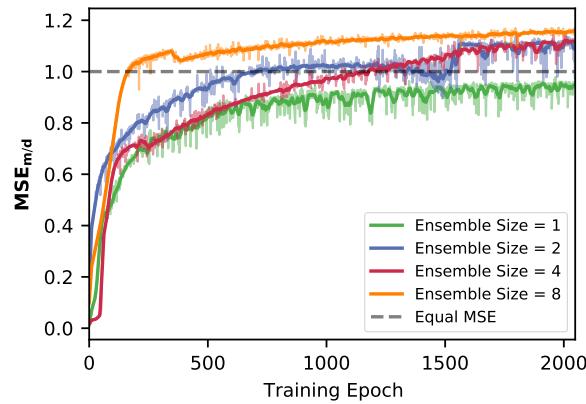


Figure 7: Joint optimization of an ensemble of articulations improves validation set image recovery performance. Bold lines represent the 16-epoch moving average of $MSE_{m/d}$, while the instantaneous values are displayed in the background.

Ensemble Size	$MSE_{m/d}(\max)$
1	0.9797
2	1.1373
4	1.1471
8	1.1713

Table 3: The maximum validation set reconstruction quality increases with ensemble size.

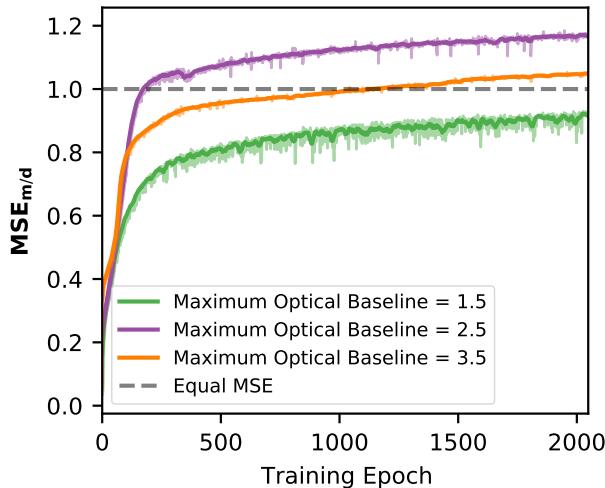


Figure 8: Relative image recovery MSE varies with maximum optical baseline (i.e., outer diameter in meters).

Outer Diameter (meters)	$MSE_{m/d}(\max)$
1.5	0.9295
2.5	1.1848
3.5	1.0535

Table 4: The maximum validation set $MSE_{m/d}$.

subapertures grow to take up a larger fraction of the pupil plane. We are currently investigating this relationship with follow-up experiments.

5. CONCLUSIONS

In this work, we introduce and examine an end-to-end differentiable model of remote sensing with distributed aperture optical systems. We show that jointly optimizing an articulation control policy and a recovery algorithm enhances the quality of reconstructed image, and confers advantages that optimizing either in isolation does not. We provide a methodology for evaluating the performance of models of this kind, and establish several baseline results upon which future work may build.

This work contains numerous compromises and simplifications which future work should address. The absence of both a focal plane model and atmospheric noise renders the problem less challenging than it will be in practice. Additionally, both the ensemble size and optical baseline experiments were limited to one run per configuration due to training resource constraints. While care was taken to ensure that the initializations of these runs were not abnormal, a systemic large-scale study should be conducted to verify our findings.

This work validates the hypothesis that task-directed apodization and learned recovery for distributed aperture systems can, in theory, provide domain-specific image recovery that matches or exceeds the quality of a monolithic design. If extended to include realistic sources of noise and validated on-sky, this approach may unlock a viable path to proliferated extended object imaging.

REFERENCES

- [1] Sydney, P. F., Africano, J. L., Fredericks, A., Hamada, K. M., Soo Hoo, V., Nishimoto, D. L., Kervin, P. W., Bisque, S., and Bisque, M., “Raven automated small telescope systems,” in [*International Symposium on Optical Science and Technology*], Bilbro, J. W., Breckinridge, J. B., Carreras, R. A., Czyzak, S. R., Eckart, M. J., Fiete, R. D., and Idell, P. S., eds., 237–247 (Oct. 2000).
- [2] Aristoff, J., Dhingra, N., Ferris, A., Hariri, A., Horwood, J., Larson, A., Lyons, T., Shaddix, J., Singh, N., and Wilson, K., “Non-traditional data collection and exploitation for improved GEO SSA via a global network of heterogeneous sensors,” in [*Proceedings of AMOS Tech Conference*], (2018).
- [3] Kuhn, J. R., Berdyugina, S. V., Langlois, M., Moretto, G., Thiébaut, E., Harlingten, C., and Halliday, D., “Looking beyond 30m-class telescopes: The Colossus project,” in [*Ground-Based and Airborne Telescopes V*], **9145**, 533–540, SPIE (July 2014).
- [4] Kuhn, J. R., Berdyugina, S. V., Capsal, J.-F., Gedig, M., Langlois, M., Moretto, G., and Thetraphihi, K., “The Exo-Life Finder Telescope (ELF): Design and beam synthesis concepts,” in [*Ground-Based and Airborne Telescopes VII*], **10700**, 344–350, SPIE (July 2018).
- [5] Kyono, T., Lucas, J., Werth, M., Calef, B., McQuaid, I., and Fletcher, J., “Machine learning for quality assessment of ground-based optical images of satellites,” *Optical Engineering* **59**, 051403 (Jan. 2020).
- [6] Kyono, T., Lucas, J., Werth, M., McQuaid, I., and Fletcher, J., “Towards Real-Time Image Reconstruction Using Deep Neural Networks,” in [*Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*], (Sept. 2020).
- [7] Lucas, J., Werth, M., Kyono, T., McQuaid, I., and Fletcher, J., “Automated Interpretability Scoring of Ground-Based Observations of LEO Objects with Deep Learning,” in [*2020 IEEE Aerospace Conference*], 1–7 (Mar. 2020).
- [8] Lucas, J., Kyono, T., Werth, M., Gagnier, N., Endsley, Z., McQuaid, I., and Fletcher, J., “Estimating Satellite Orientation through Turbulence with Deep Learning,” in [*Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*], (Sept. 2020).
- [9] Yang, J., Lucas, J., Kyono, T., Abercrombie, M., Fletcher, J., and McQuaid, I., “Semantic Segmentation of Low Earth Orbit Satellites using Convolutional Neural Networks,” AMOS (2021).
- [10] Tseng, E., Colburn, S., Whitehead, J., Huang, L., Baek, S.-H., Majumdar, A., and Heide, F., “Neural nano-optics for high-quality thin lens imaging,” *Nature Communications* **12**, 6493 (Nov. 2021).
- [11] Ramsundar, B., Krishnamurthy, D., and Viswanathan, V., “Differentiable Physics: A Position Piece,” *arXiv:2109.07573 [physics]* (Sept. 2021).
- [12] Yuan, Q., Wei, Y., Meng, X., Shen, H., and Zhang, L., “A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **11**(3), 978–989 (2018).
- [13] He, K., Zhang, X., Ren, S., and Sun, J., “Deep residual learning for image recognition,” in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 770–778 (2016).
- [14] Ronneberger, O., Fischer, P., and Brox, T., “U-net: Convolutional networks for biomedical image segmentation,” in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 234–241, Springer (2015).
- [15] Jenkins, F. A. and White, H. E., “Fundamentals of optics,” *Indian Journal of Physics* **25**, 265–266 (1957).

- [16] Steward, E. G., [*Fourier Optics: An Introduction*], Courier Corporation (2004).
- [17] Kingma, D. P. and Ba, J., “Adam: A Method for Stochastic Optimization,” *arXiv:1412.6980 [cs]* (Jan. 2017).
- [18] Park, T. H., Märtens, M., Lecuyer, G., Izzo, D., and D’Amico, S., “SPEED+: Next-Generation Dataset for Spacecraft Pose Estimation across Domain Gap,” *arXiv:2110.03101 [cs]* (Dec. 2021).
- [19] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., and Isard, M., “Tensorflow: A system for large-scale machine learning,” in [*12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*], 265–283 (2016).